



The Three Principles of Data Center Infrastructure Design

Tuan Nguyen, Sunset Learning Institute Cisco Specialized Instructor

Your data center is the most critical resource within your organization. As you know, it provides the means for all storage, management and dissemination of data, applications and communications for your business. When employees and customers are unable to access the servers, storage systems and networking devices that reside in the data center, your entire organization can shut down. Millions of dollars can be lost in a matter of minutes for businesses like banks, airlines, shipping facilities and online brokerages. Faced with these consequences, IT executives today must optimize their data centers, particularly the network infrastructure. When you consider that 70 percent of network downtime can be attributed to physical layer problems, specifically cabling faults, it's paramount that more consideration be given to infrastructure design.

The data center is home to the computational power, storage, and applications necessary to support an enterprise business. The data center infrastructure is central to the IT architecture, from which all content is sourced or passes through. Proper planning of the data center infrastructure design is critical, and performance, resiliency, and scalability need to be carefully considered.

Another important aspect of the data center design is flexibility in quickly deploying and supporting new services. Designing a flexible architecture that has the ability to support new applications in a short time frame can result in a significant competitive advantage. Such a design requires solid initial planning and thoughtful consideration in the areas of port density, access layer uplink bandwidth, true server capacity, and oversubscription, to name just a few.

The data center network design is based on a proven *layered* approach, which has been tested and improved over the past several years in some of the largest data center implementations in the world. The layered approach is the basic foundation of the data center design that seeks to improve scalability, performance, flexibility, resiliency, and maintenance.

The layers of the data center design are the *core*, *aggregation*, and *access* layers. These layers are briefly described as follows:

- ✓ **Core layer**—Provides the high-speed packet switching backplane for all flows going in and out of the data center. The core layer provides connectivity to multiple aggregation modules and provides a resilient Layer 3 routed fabric with no single point of failure. The core layer runs an interior routing protocol, such as OSPF or ISIS, and load balances traffic between the campus core and aggregation layers.
- ✓ **Aggregation layer modules**—Provide important functions, such as service module integration, Layer 2 domain definitions, spanning tree processing, and default gateway redundancy. Server-to-server multi-tier traffic flows through the aggregation layer and can use services, such as firewall and server load balancing, to optimize and secure applications.
- ✓ **Access layer**—Where the servers physically attach to the network. The server components consist of 1RU servers, blade servers with integral switches, blade servers with pass-through cabling, clustered servers, and mainframes with OSA adapters. The access layer network infrastructure consists of modular switches, fixed configuration 1 or 2RU switches, and integral blade server switches. Switches provide both Layer 2 and Layer 3 topologies, fulfilling the various servers broadcast domain or administrative requirements.



Principle 1: Space Savings

Environmentally controlled real estate is expensive. Data center racks and equipment can take up an enormous amount of real estate, and the future demand for more network connections, bandwidth and storage may require even more space. With insufficient floor space as the topmost concern among IT managers today, maximizing space resources is the most critical aspect of data center design. Business environments are constantly evolving, and as a result, data center requirements continuously change. Providing plenty of empty floor space when designing your data center enables the flexibility of reallocating space to a particular function, and adding new racks and equipment as needed.

Principle 2: Reliability

Uninterrupted service and continuous access are critical to the daily operation and productivity of your business. With downtime translating directly to loss of income, data centers must be designed for redundant, fail-safe reliability and availability. Data center reliability is also defined by the performance of the infrastructure. As information is sent back and forth within your facility and with the outside world, huge streams of data are transferred to and from equipment areas at extremely high data rates. The infrastructure must consistently support the flow of data without errors that cause retransmission and delays. As networks expand and bandwidth demands increase, the data center infrastructure must be able to maintain constant reliability and performance.

Principle 3: Manageability

Manageability is key to optimizing your data center. The infrastructure should be designed as a highly reliable and flexible utility to accommodate disaster recovery, upgrades and modifications. Manageability starts with strategic, unified cable management that keeps cabling and connections properly stored and organized, easy to locate and access, and simple to reconfigure.

Cable routing paths must be clearly defined and intuitive to follow while enabling easy deployment, separation, access, reduced congestion, and room for growth. This is especially important in data centers with large volumes of cables. Cables managed in this way improve network reliability by reducing the possibility of cable damage, bend radius violations, and the time required for identifying, routing, and rerouting cables.

The use of a central patching location in a cross-connect scenario provides a logical and easy-to-manage infrastructure whereby all network elements have permanent equipment cable connections that once terminated, are never handled again. The advantage to deploying centralized patching in your data center includes:

- Lower operating costs by greatly reducing the time it takes for modifications, upgrades and maintenance.
- Enhanced reliability by making changes on the patching field rather than moving sensitive equipment connections.
- Reduced risk of down time with the ability to isolate network segments for troubleshooting and quickly reroute circuits in a disaster recovery situation.